

TEST DE RAVEN, BAREMOS ARGENTINOS PARA EL RANGO 13-18 AÑOS Y EFECTO FLYNN

*Lilia Rossi-Casé, Rosa Neer, Susana Lopetegui, Stella Doná,
Bruno Biganzoli & Ramiro Garzaniti**

Resumen

A través de este estudio se muestran los resultados obtenidos de la administración Test de Raven, Escala General, sobre una muestra de 1067 estudiantes argentinos entre 13 y 18 años. La cantidad de casos es proporcional al de la población (INDEC, 2010). Para su análisis, se establecieron tres grupos de edad con intervalos iguales de 2 años. Las distribuciones de frecuencias tienden a ser normales en los tres casos. Las puntuaciones directas observadas aumentan a medida que aumenta la edad y la variabilidad de las mismas, disminuye. Quince años después, se corrobora la vigencia del efecto Flynn constatada en el año 2000 (Rossi-Casé, Neer, y Lopetegui, 2001). A ello se añade la verificación de una estabilidad o *amesetamiento* de los puntajes promedios alcanzados a partir de esa fecha, para los grupos 15-16 años y 17-18 años, y una disminución no significativa para los puntajes de los sujetos de la franja de 13-14 años ($p < .05$) respecto de los baremos del año 2000. Asimismo, se observa una alteración en la progresión de la dificultad entre las series C y D. Todo ello en consonancia con investigaciones similares en distintas partes del mundo.

Palabras clave: Test de Raven, baremos, 13-18 años, Efecto Flynn

* Licenciados en Psicología, docentes de la Facultad de Psicología; Instituto de Investigaciones en Psicología (InIPsi) – UNLP. E-mail: rossicase@psico.unlp.edu.ar.

TESTE DE RAVEN, TABELAS ARGENTINAS PARA O RANGO 13-18 ANOS E EFEITO FLYNN

A través de este estudio se muestran los resultados obtenidos de la administración del Teste de Raven, Escala Geral, sobre una muestra de 1067 estudiantes argentinos entre 13 y 18 años. La cantidad de casos es proporcional a la de la población (INDEC, 2010). Para su análisis, se establecieron tres grupos de edad con intervalos iguales de 2 años. Las distribuciones de frecuencias tienden a ser normales en los tres casos. Las puntuaciones directas observadas crecen a medida que crece la edad y disminuye la variabilidad de las mismas. Quince años después, corrobórase la vigencia del efecto Flynn constatada en el año 2000 (Rossi-Casé, Neer, y Lopetegui, 2001). Agrega-se a la verificación de una estabilidad o “platô” de las puntuaciones medias alcanzadas a partir de ese dato, para los grupos 15-16 años y 17-18 años, y una disminución no significativa para las puntuaciones de los sujetos de la franja de 13-14 años ($p < .05$) respecto de las tablas del año 2000. También se observa un cambio en la progresión de la dificultad entre las series C y D. Todo esto en consonancia con investigaciones similares en distintas partes del mundo.

Palabras clave: Teste de Raven, tablas, 13-18 años, Efecto Flynn

Introducción

En este trabajo se presentan los resultados del Proyecto de Investigación “*Teste de Raven. Escala General. Normas Percentilares para las edades 13 a 18 años para la ciudad de La Plata*”, acreditado por la Facultad de Psicología de la Universidad Nacional de La Plata, desarrollado entre los años 2012-2015.

Las pruebas de evaluación psicológica integran el conjunto de instrumentos que permiten al profesional psicólogo arribar a un diagnóstico respecto de aquellos sujetos sobre los que recae la investigación. Estas pruebas aportan información crucial para el

desarrollo de estrategias de intervención específicas, tanto para el diagnóstico como para el pronóstico. Las puntuaciones directas que obtienen los sujetos son transformadas a baremos que permiten su comparación con una muestra estandarizada y dan un sentido estadístico al desempeño individual. Pero para que la interpretación diagnóstica sea correcta los baremos deben estar actualizados, es decir, el grupo de referencia del cual se obtienen las puntuaciones promedio con las que se compara el rendimiento de un sujeto, debe ser el adecuado. Asimismo, la revisión periódica de los tests en función de los nuevos descubrimientos e hipótesis que aportan las teorías así como los cambios culturales y las adaptaciones regionales, se vuelve necesaria para que el test aplicado sea válido y confiable (Casullo, 2009).

Un aspecto especial a ser considerado es que los puntajes medios obtenidos en los tests de inteligencia han aumentado regularmente y de manera notable en todo el mundo. Este fenómeno se observa desde la misma aparición de los tests psicométricos. James R. Flynn ha sintetizado los resultados de las investigaciones sobre el tema (Flynn, 1984) y, a raíz de ello, al aumento constante de los valores absolutos en todos los tests de inteligencia se denomina *Efecto Flynn*. Lo mencionado hasta aquí advierte que las normas de los tests psicológicos en modo alguno son absolutas, universales o permanentes. Al respecto, Casullo (2009) resalta la necesidad de plantearnos universales al proponernos estudiar una porción de la conducta a través de una técnica de evaluación psicológica; pero al momento de construirla, esta debe respetar la particularidad de la cultura donde va a ser aplicada.

El presente proyecto buscó obtener datos actualizados y compararlos con aquellos de las estandarizaciones de los años 1964 y 2000, para muestras de iguales características con el propósito de verificar la presencia o no del efecto Flynn en esta población.

Por otro lado, dada la importancia del Test de Matrices Progresivas de Raven (Raven, J. C., Raven y Court, 2003) como instrumento de evaluación psicológica y su uso generalizado con diferentes propósitos en distintos ámbitos de aplicación de la Psicología, se hace necesaria la actualización de sus baremos para poder utilizarlo de forma confiable.

Test de Matrices Progresivas de Raven

El Test de Raven fue publicado por primera vez en el año 1938 por su autor, John C. Raven, alumno del psicólogo inglés Charles Spearman, quien había enunciado en 1904 su *“teoría ecléctica de los dos factores”*. Se trata de un test no verbal, de capacidad intelectual, de habilidad mental general. Examina el factor G, para lo cual pone en juego procesos de educación de relaciones y correlaciones sobre un material en el que las variables no son obvias, es decir, que se deben extraer nuevas comprensiones a partir de la información dada.

Su administración puede ser individual o colectiva, es autoadministrable, sin límite de tiempo en su versión original. Es un test de matrices lacunarias, en el que la tarea del sujeto consiste en completar una matriz, debiendo elegir la respuesta correcta de entre seis u ocho alternativas. Se presenta en varias formas: Escala General, Escala Paralela, Escala Especial Coloreada y Escalas Avanzadas (Raven, J. C. et al., 2003). La prueba fue revisada en sucesivas oportunidades, generalmente con el fin de obtener normas o baremos adecuados a diferentes poblaciones.

La mencionada capacidad eductiva se define como la capacidad de extraer nuevas comprensiones e información partiendo de lo que se percibe o ya es conocido (Raven, J. C. et al., 2003). La educación, como operación cognoscitiva básica, se entiende a partir de las tres leyes en que Spearman descompone el proceso de formación del

conocimiento, conocidas como *leyes no genéticas*. La primera de ellas, que indica que ante dos o más ítems toda persona tiende a establecer relaciones entre ellos, se denomina ley de *educación de relaciones* (Raven, J. C., Raven, J. et al., 2003). La segunda, llamada de *educación de correlatos*, establece que ante un ítem y una relación, toda persona tiende a concebir el ítem correlativo. La tercera, según la cual toda persona tiende a conocerse de un modo inmediato a sí misma y a los ítems de su propia experiencia, se conoce como *ley de autoconciencia o introspección*. Entonces, la conducta educativa exige un proceso perceptivo más activo que uno analítico o reproductivo; supone problematizar lo familiar, así como desarrollar la comprensión y resolver problemas. Conceptualmente estaría relacionada con la *inteligencia fluida*, mientras que la capacidad reproductiva se aproxima a la *inteligencia cristalizada* (Cattell, 1968). Inteligencia fluida refiere a la habilidad para utilizar el razonamiento abstracto para resolver nuevos problemas que no fueron enseñados (Kaufman, 2009). Generalmente se evalúa mediante analogías abstractas cuyos datos no envejecen ni son específicos de una cultura. La inteligencia cristalizada es culturalmente específica, dependiente de la escolarización formal y sujeta a los conocimientos adquiridos, por ende, su incremento a lo largo de la vida es esperable, mientras que el incremento de la inteligencia fluida no (Kaufman, 2009).

El Test de Raven es un test que mide la inteligencia fluida, por lo que coincidimos con Fernández Liporace, Ongarato, Saavedra y Casullo (2004) al afirmar que “*esto explica la utilización tan difundida de las Matrices en los ámbitos de investigación*”, ya que los resultados que aporta permiten comparar poblaciones y/o sujetos que “*han sido expuestos a situaciones educativas formales e informales de naturaleza bien disímil*” (pp. 50-69).

Debido a la constatación del aumento intergeneracional en los niveles de capacidad educativa, se alcanzó un “efecto techo” en los adolescentes y adultos jóvenes más capaces, siendo cada vez más frecuente encontrar casos donde el puntaje directo del test es el máximo posible. Esto introdujo, tras comprobarlo en las estandarizaciones británicas del Test de Raven de 1979 en jóvenes, y de 1992 en adultos (Raven, J. C. et al., 2003), la necesidad de extender el rango de dificultad de los ítems para devolver al test la capacidad de discriminación que originalmente tenía en estos grupos de sujetos. Así, en la revisión del test de 1998 (Raven, J., Raven y Court, 2003) se incluyeron las formas Paralela y Plus de la Escala General del Test.

La Escala General Plus (Raven, J. et al., 2003), que conserva el formato cíclico en 60 ítems de la Escala General Clásica, satisfizo la necesidad de ampliar el rango de dificultad de los ítems sin reducir el poder de discriminación del test entre los sujetos menos capaces (Raven, J. et al., 2003). Asimismo, para los casos en que se necesita una discriminación aún mayor en los extremos inferior y superior de la distribución, se cuenta con la posibilidad de recurrir a las Escalas Coloreada y Avanzada, respectivamente (Raven, J. C. et al., 2003).

Simultáneamente se desarrolló también una versión Paralela (Raven, J. et al., 2003), que al corresponderse con la versión Clásica en forma general e ítem por ítem en cuanto a estrategias de solución y dificultad empírica, puede utilizarse con los baremos ya existentes para la Escala General, de manera tal que la forma original y la forma paralela resultan distintas pero equivalentes. La creación de esta última versión responde a la necesidad de salvar las dificultades generadas por la popularidad del test, pues al ser demasiado conocido podría ocurrir que algunas personas se entrenaran para lograr mejores resultados. Sobre este tema, Rossi-Casé, Doná y Garzaniti (en prensa, 2015) realizaron un estudio piloto descriptivo correlacional sobre una muestra de 232

sujetos de la ciudad de La Plata, Argentina. El coeficiente de correlación de Pearson fue de $r = 0.728$, con valores promedios y desviaciones típicas similares para ambas escalas.

El efecto Flynn

A partir del uso generalizado de los tests de inteligencia, se ha observado que los puntajes medios obtenidos aumentan regularmente y de manera notable con el correr del tiempo *en todo el mundo*. Esta constatación indica que, para la misma población, las normas para el cociente intelectual se vuelven obsoletas conforme pasan los años. Es decir que, en la actualidad, para obtener el mismo puntaje transformado que hace algunas décadas, las puntuaciones directas deberían ser más altas, debiendo resolverse un mayor número de problemas. Estos aumentos son mayores en aquellos tests que miden la inteligencia fluida, que en los que miden la inteligencia cristalizada (Sundet, Barlaug y Torjussen, 2004). Esto llevó a pensar que el aumento está relacionado al factor G de la inteligencia y no a los factores específicos (i.e., factores E). La causa de este fenómeno permanece desconocida y no existe evidencia suficiente que permita afirmar que este hecho refleje un verdadero aumento de la inteligencia, ligado posiblemente a factores tales como la heterosis, la mejora en las condiciones de vida en las distintas poblaciones, alimentación o nutrición más adecuada; la expansión del sistema educativo; la reducción del tamaño de las familias nucleares, con el consecuente incremento del desarrollo psicológico; la adquisición progresiva de ciertas habilidades para responder satisfactoriamente a los tests; un mayor desarrollo de la escolarización y educación de los niños; la creciente tecnologización de la cultura, desde los juegos de video hasta el acceso cada vez más irrestricto a los medios de comunicación a través de internet, que ofrecen otros aspectos de estimulación (Sundet et al., 2004). Por su parte, Armstrong y Woodley (2014) afirman que hay evidencia neurológica que señala que el

efecto Flynn está asociado a un aumento en el tamaño del cerebro que ha potenciado las funciones del hipocampo, aunque no descartan que otra causa para este fenómeno sea la posibilidad de que los tests que miden la inteligencia fluida y la capacidad eductiva, como el Test de Matrices Progresivas de Raven, generen algún tipo de *andamio cognitivo* que permita el uso de habilidades que no dependan del factor G para resolver los problemas.

En su estudio inicial, Flynn (1984) estableció que la magnitud del aumento del cociente intelectual, medido inicialmente con las escalas Wechsler y Stanford-Binet, era de 0.3 puntos por año o 3 puntos por década. Estos estudios se realizaron comparando muestras estadounidenses entre los años 1932 y 1978. Más adelante, la misma tendencia se corroboró con datos de otros 20 países, incluyendo Argentina (Rossi Casé et al, 2002).

No obstante, investigaciones recientes muestran una disminución en el crecimiento de dichos puntajes. Este efecto fue observado por primera vez en los países escandinavos. Tal es el caso de investigaciones longitudinales hechas en Noruega (Sundet et al., 2004) en las cuales se muestra una disminución en los puntajes de una batería de tests administrados a los jóvenes de 18 años, antes de entrar al servicio militar, desde los años 50. Las pruebas utilizadas son tests de matemática y lenguaje, similares a los subtests del WAIS IV, y un test no verbal que fue construido de forma similar al test de Raven. Los dos primeros miden la inteligencia cristalizada, mientras que el último mide la inteligencia fluida. Teasdale y Owen (2007) estudiaron los datos aportados por una batería similar en Dinamarca, la cual desde 1957 se administra a todos los jóvenes de 18 años que ingresan al servicio militar. De los cuatro tests que la componen, los autores afirman que el de Matrices de Letras se asemeja a las Matrices Progresivas de Raven. Los autores encontraron que si bien hubo un pequeño incremento de los puntajes entre

1988 y 1998, los mismos disminuyeron en 2003-2004, incluso por debajo de los obtenidos en 1988 para todos los tests de la batería. Esto se observó en jóvenes de todos los niveles educativos. Con esto, los autores apoyan la afirmación de que en lo que va del siglo XXI ha habido muy poca evidencia sobre la continuidad del efecto Flynn, tal como se lo define hasta este momento, es decir, como el aumento continuo de las medias en los tests de inteligencia.

Por su parte, Brouwers, Van de Vijver y Van Hemert (2008) hicieron una revisión de 193 estudios sobre las tres escalas Raven publicados entre 1944 y 2003, que abarcan 798 muestras provenientes de 45 países, siendo el número total de sujetos 244316. Dentro de ese reporte, Argentina se encuentra representada por más de 20 estudios publicados en dicho período. Los autores calcularon la correlación entre las características de las muestras (edad, años de escolaridad), de los países de los que provenía cada estudio (Producto Bruto Interno [PBI], escolaridad y analfabetismo), y de los estudios consultados (año de publicación). Encontraron que, si bien tanto el PBI como los años de escolaridad se correlacionan positivamente con el desempeño en el test (por lo que pueden estar influyendo en las diferencias interculturales en el desempeño), existe una correlación negativa significativa entre el PBI y la magnitud del efecto Flynn. A partir de esto, concluyen que el efecto Flynn habría alcanzado su techo en los países desarrollados de occidente, mientras que los países con un PBI per cápita menor, todavía muestran un efecto Flynn más pronunciado. Esta conclusión coincide con la de Teasdale y Owen (2007), quienes afirman que si el efecto Flynn llega a su fin en los países más desarrollados, estaría lejos de hacerlo en el resto del mundo, pero que se esperaría que estas diferencias entre los distintos países disminuyeran gradualmente en el futuro. Flynn (2013) ha resaltado que también Holanda y Finlandia muestran una disminución en la magnitud del efecto que lleva su nombre, sin embargo, esta tendencia

no se evidencia en todos los países desarrollados. Los datos recientes provenientes de Australia son ambiguos al respecto y Estados Unidos y Corea del Sur constituyen claras excepciones a la misma. También menciona que algunos autores atribuyeron esta tendencia a la inmigración de poblaciones con CI bajos en los países desarrollados, pero que esto no explicaría la situación de Estados Unidos. Flynn aventura la idea de que la detención de los incrementos del CI en ciertos países podría deberse a que en ellos ya no puede progresarse más en aquellos aspectos que se supone fomentarían dichos incrementos. Rusell (2007) sostiene que un efecto meseta sería de esperar en culturas que tengan condiciones de vida óptimas, lo cual explicaría que dicho efecto se haya manifestado en los países escandinavos antes que en el resto del mundo, ya que estos han establecido un sistema de bienestar que incluye a toda la población desde el fin de la Segunda Guerra Mundial. Recurriendo a las escalas Wechsler y empleando la regresión lineal, este autor estima que en Estados Unidos el efecto meseta se alcanzará en 2024. Pero al considerar posibles sesgos en la selección de las muestras para las estandarizaciones previas de estas escalas, estima que la meseta ya podría haberse alcanzado en 2004, por lo que el efecto Flynn podría ya no estar vigente en ese país.

Con respecto a la Argentina, este equipo realizó una primera constatación del efecto Flynn en la ciudad de La Plata y alrededores al actualizar los baremos del Test de Raven en el año 2000, y compararlos con la estandarización del año 1964. Esta comparación mostró un considerable aumento en los puntajes directos del test para todas las edades (Flynn y Rossi-Casé, 2012; Rossi-Casé, Neer y Lopetegui, 2001, 2002, 2011; Rossi-Casé et al., 2014). Los resultados sirvieron para comprobar tendencias a nivel mundial (Flynn y Rossi-Casé, 2011).

Método

Participantes

La muestra de estandarización se conformó con 1067 sujetos de ambos sexos, con edades entre los 13 y los 18 años cumplidos al momento de la administración del test. Se dividió ese rango, en tres intervalos iguales de dos años cada uno. Ello se debe a que el rendimiento en la prueba cambia de modo atendible según la edad, razón por la cual se justifica la generación de normas diferentes para cada grupo de edades (Casullo, 2009).

La muestra fue estratificada teniendo en cuenta las variables edad, sexo y tipo de gestión de la institución educativa (de gestión pública o privada), de alumnos de educación secundaria y universitaria; de La Plata, Argentina.

Para la extracción de la muestra, se tuvieron en cuenta los datos del censo poblacional del año 2010 (Instituto Nacional de Estadística y Censos [INDEC], 2010). Se utilizó un muestreo bietápico que permitió seleccionar distintos conglomerados según las instituciones educativas y, en cada una de ellas, elegir aleatoriamente los grupos de clase que serían examinados para las distintas edades, siendo evaluados todos los sujetos que conformaban cada grupo.

Se trabajó con un intervalo de confianza del 95% que permitió establecer que el error muestral máximo sería del 3% para este tamaño de muestra. En la Tabla 1, se describe la composición de la muestra, según sexo y edad. En la Tabla 2 se comparan los porcentajes de la composición poblacional y muestral, por rango de edades.

(insertar Tablas 1 y 2)

Instrumentos

Se utilizó el Test de Matrices Progresivas de Raven; Escala General, segunda edición (Raven, Raven, & Court, 2003). El test está compuesto por 60 problemas organizados

en cinco series (A, B, C, D y E), de doce ítems cada uno. En cada una de las series los ítems están ordenados por dificultad creciente, comenzando con problemas sencillos. Cada serie resulta con un grado de dificultad mayor que la anterior. La dificultad creciente entre series es también dentro de cada una de las cinco series de modo tal que el último problema de una serie resulta más difícil que el primero de la serie inmediata posterior.

Procedimientos

El test se administró de manera colectiva, en grupos-clase, con presencia del examinador y sin límite de tiempo para su ejecución, para permitir la evaluación de la capacidad intelectual sin la intervención de la velocidad para resolver la tarea.

Para la participación de los sujetos de la muestra se solicitó, en tiempo y forma, la debida autorización a los Inspectores Educativos del Distrito La Plata y se obtuvo el consentimiento informado previo de los padres o tutores de los adolescentes, mediante un formulario donde se explicaba el propósito de la investigación y la confidencialidad de los datos. En el caso de los alumnos con mayoría de edad jurídica, al cumplir los 18 años de edad para el caso de la Argentina, se requirió el consentimiento de manera directa a cada uno de ellos.

El test fue administrado en distintos períodos de los años 2012-2013 y 2014-2015 (noviembre de 2012; junio, octubre y noviembre de 2013; mayo y junio de 2014; marzo, abril y mayo de 2015).

La administración estuvo a cargo de profesionales y alumnos avanzados de la carrera de Psicología, entrenados previamente a fin de homogeneizar el procedimiento.

Para la consigna, se siguieron de manera rigurosa las instrucciones dadas por el autor del test para la administración colectiva. La administración se llevó a cabo en cada una de las instituciones educativas seleccionadas, en horarios regulares de clases, cedidos

por las autoridades y profesores de cada institución. Cada grupo no superó los 36 alumnos.

Al conformar la base de datos, se excluyeron aquellos protocolos que mostraban una discrepancia mayor a la esperada en la composición de los puntajes ($n = 148$). Finalmente se analizó una base de datos que incluye 1067 protocolos.

Análisis

Con la información obtenida se conformó una base de datos (Programa Excel, Microsoft, 2007) obteniéndose los estadísticos descriptivos correspondientes así como los percentiles que permitieron elaborar las normas para este grupo.

Los resultados obtenidos en el estudio actual se compararon con las normas obtenidas en los años 1964 y 2000 para los mismos grupos etarios de la ciudad de La Plata, Argentina, utilizando el estadístico t de Student con un nivel de confianza de 95%.

El análisis de datos se realizó por grupos de edad: 13-14 años; 15-16 años y 16-17 años, para poder comparar los valores percentilares con otros estudios de la misma índole. Para los mismos grupos se compararon además los puntajes obtenidos en las series para poder estimar su grado de dificultad.

Resultados

La descripción de las puntuaciones directas obtenidas muestra que en los tres grupos de edad el rango de respuestas correctas varía de 10 a 60. Las puntuaciones mínimas resultaron similares para todas las edades. La puntuación máxima posible de 60, es alcanzada en los tres grupos. Los resultados expuestos en la Tabla 3 muestran que, a medida que aumenta la edad de los sujetos, el rendimiento de los grupos mejora y la heterogeneidad de sus respuestas disminuye.

(insertar Tabla 3)

La comparación a través de los años, de las puntuaciones directas necesarias para que un sujeto se ubique el percentil 50, se presenta en la Tabla 4.

(insertar Tabla 4)

El análisis de las puntuaciones promedios de cada Serie, por rango de edades, muestra que el orden creciente de dificultad para la resolución de los ítems del Test se altera en el caso de las series C y D. En todos los grupos de edad estudiados los puntajes de la serie C resultaron menores a los de la serie D.

(insertar Tabla 5)

Resultados para el grupo de edad 13-14 años

En los sujetos que integran este grupo se observan los resultados más bajos de todos los que integran la muestra. La puntuación promedio es de 44.05 puntos y la desviación estándar es de 7.60 puntos. El rango de respuestas correctas fue de 12 a 60.

La distribución de las puntuaciones directas necesarias para la ubicación de los sujetos en la escala percentilar se muestra en la Figura 1. Los datos tienden a ajustarse a la distribución normal.

(insertar Figura 1)

La mitad de los sujetos obtuvo 45 puntos o menos. Si bien esto representa un incremento de 6 puntos respecto del baremo del año 1964, el resultado observado actualmente es menor en 3 puntos que el alcanzado en el año 2000.

Este incremento significativo de las puntuaciones respecto de las normas 1964 ($t(359)=1.98$; $p < .05$) y la disminución respecto de los baremos del año 2000, se observó en todos los valores percentilares calculados. Estas observaciones se aprecian en la Figura 2.

(insertar Figura 2)

Resultados para el grupo de edad 15-16 años

La puntuación promedio observada es de 48.39 puntos y la desviación estándar es de 6.75 puntos. El rango de respuestas correctas fue de 10 a 60.

La distribución de las puntuaciones directas necesarias para la ubicación de los sujetos en la escala percentilar se muestra en la Figura 3. Los datos tienden a ajustarse a la distribución normal.

(insertar Figura 3)

La mitad de los sujetos obtuvo al menos 50 puntos. Esto representa un incremento de 9 puntos respecto del baremo del año 1964, y de 1 punto respecto de la norma alcanzada en el año 2000.

Este incremento de las puntuaciones respecto de las normas 1964 se observó en todos los valores percentilares calculados ($t(334)=1.41$; $p < .05$). La comparación de resultados con los baremos del año 2000 muestra aumento de 1 y 2 puntos para los percentiles 25, 50 y 75, mantenimiento de los puntajes para los percentiles 10, 90 y 95; y una disminución de 1 punto para percentiles 5 y 99. Estas observaciones se aprecian en la Figura 4.

(insertar Figura 4)

Resultados para el grupo de edad 17-18 años

En los sujetos que integran este grupo se observan los resultados más altos de todos los que integran la muestra. La puntuación promedio es de 49.23 puntos y la desviación estándar de 6.39 puntos. El rango de respuestas correctas fue de 12 a 60.

La distribución de las puntuaciones directas necesarias para la ubicación de los sujetos en la escala percentilar se muestra en la Figura 5. Los datos tienden a ajustarse a la distribución normal.

(insertar Figura 5)

La mitad de los sujetos obtuvo 50 puntos o menos. El resultado observado representa un incremento de 10 puntos respecto del baremo del año 1964, al tiempo que permanece igual a la norma alcanzada en el año 2000. Este incremento significativo de las puntuaciones respecto de las normas 1964 ($t(370)=0.64$; $p < .05$) se observó en todos los valores percentilares calculados. Respecto de los baremos del año 2000, se constata la disminución de 3 y 2 puntos para los valores percentiles 5 y 10; igualdad en las puntuaciones de los percentiles 25, 50, 90 y 95; y un aumento de un 1 punto en los valores de los percentiles 75 y 99. Estas observaciones se aprecian en la Figura 6.

(insertar Figura 6)

Discusión y conclusiones

El efecto Flynn corroborado en el año 2000 respecto de las normas del año 1964 (Rossi-Casé, Neer y Lopetegui, 2001), no se observa entre el baremo del año 2000 y el estudio actual.

Concordando con investigaciones recientes realizadas en otros países (Brouwers, Van de Vijver y Van Hemert, 2008; Rusell, 2007; Sundet et al., 2004; Teasdale y Owen, 2007), los resultados obtenidos permiten mostrar una detención del aumento de los puntajes directos necesarios para alcanzar el rendimiento promedio en cada rango de edad.

En el estudio actual se constata una leve disminución de las puntuaciones promedio para el rango de edad 13-14 años que no resulta significativa, y una paridad para los otros dos grupos de edad. Es en este sentido que hablamos de un *amesetamiento* de los puntajes.

El efecto Flynn evidencia que los puntajes han ido creciendo a razón de más o menos una desviación estándar por generación (Flynn, 1984). Resulta adecuado preguntarse si

este aumento será indefinido a lo largo del tiempo o por el contrario en algún momento cesará y en tal caso cuando.

Para intentar dar respuesta a esta cuestión deberíamos volver a analizar las hipótesis que en su momento intentaron explicar la presencia del efecto Flynn, tales como mejoras en la alimentación, mayor escolarización, el uso de las nuevas tecnologías y el predominio de las imágenes visuales, entre otras.

Por lo tanto, si las condiciones que alguna vez se mencionaron para explicar la diferencia se han estabilizado en los últimos años, los puntajes, que son su expresión, han corrido la misma suerte.

No hay evidencia de que exista alguna otra condición que produzca un cambio cualitativo de aquel tenor; entonces, las puntuaciones directas no tienen ahora por qué cambiar de manera significativa.

Esto puede llevar a entender el *amesetamiento* observado en Argentina y otros países del mundo.

Otra explicación de este fenómeno podría ser que el Test de Matrices Progresivas de Raven no estuviese midiendo únicamente lo que dice medir, y que otros factores fueran los que incidan en su resolución, como por ejemplo la educación formal y, en consecuencia, la inteligencia cristalizada.

Estos factores podrían ser más sensibles a los estímulos del ambiente, lo que podría dar cuenta no sólo de los cambios que aparecen a lo largo de los años, sino también las diferencias que existen entre los países desarrollados y aquellos en vías de desarrollo.

Hay un aspecto llamativo que surge del análisis de los resultados obtenidos en este estudio y es que no se cumple con la progresión de dificultad en la resolución de las matrices de las series C y D.

Para los tres grupos de edad estudiados, el promedio de las puntuaciones directas de la serie C mostró ser más bajo que las de la serie D, resultado contrario a lo esperado según el diseño del Test de Raven.

Esta graduación puede incidir sobre los sujetos examinados, en la medida en que deben esperar mayor dificultad cuando las matrices se van sucediendo y no en sentido contrario, tal como se observa en el presente estudio.

Los mismos coinciden con lo observado en investigaciones realizadas por Fernández-Liporace (2004). En ambos casos se sugiere la necesidad de examinar con mayor profundidad estos resultados con el propósito de reestructurar las series C y D para devolver a la prueba su original diseño de dificultad creciente.

Enviado: 14-8-2015
Revisión recibida: 12-9-2015
Aceptado: 20-10-2015

Referencias

Armstrong, E. L. y Woodley, M. A. (2014). The rule-dependence model explains the commonalities between the Flynn effect and IQ gains via retesting. *Learning and Individual Differences*, 29, 41-49. Recuperado de <http://dx.doi.org/10.1016/j.lindif.2013.10.009>

Brouwers, S. A., Van de Vijver, F. J. R. y Van Hemert, D. A. (2008). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences*, 19, 330-338. Recuperado de <http://dx.doi.org/10.1016/j.lindif.2008.10.006>

Casullo, M. M. (2009). La evaluación psicológica: modelos, técnicas y contextos. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 27(1), 9-28.

Cattel, R. B. (1968). The theory of fluid and crystallized intelligence. Its relationship to culture free tests and its verification in 9-12 ys. old children. *Bolletino di Psicologia Applicato*, 88(90), 3-22.

Fernández-Liporace, M., Ongarato, P., Saavedra, E. y Casullo, M. M. (2004). El Test de Matrices Progresivas, Escala General: un análisis psicométrico. *Evaluar*, 4, 50-69.

Flynn, J. R. (1984). The mean IQ of Americans: massive gains 1932-1978. *Psychological Bulletin*, 95, 29-51.

Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence*, 41(6), 851-857. doi:10.1016/j.intell.2013.06.014

Flynn, J. R. y Rossi-Casé, L. (2011). Modern women match men on Raven's Progressive Matrices. *Personality and Individual Differences*, 50(6), 799-803. <http://dx.doi.org/10.1016/j.paid.2010.12.035>

Flynn, J. R. y Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, 40, 145-150. Recuperado de <http://dx.doi.org/10.1016/j.intell.2012.01.006>

Instituto Nacional de Estadística y Censos. (2010). *Censo Nacional de Población, Hogares y Viviendas 2010*. Recuperado de <http://www.censo2010.indec.gov.ar/>

Kaufman, A. S. (2009). Hot Topic: Are Our IQs Fixed or Are They Malleable? En J. C. Kaufman (Ed.), *IQ Testing 101*, (pp. 201-222). New York: Springer.

Raven J. C., Raven, J. y Court, J. H. (2003). *Test de Matrices Progresivas: Escala General*. Buenos Aires: Paidós.

Rossi-Casé, L., Neer, R. y Lopetegui, S. (2001). Test de Matrices Progresivas de Raven: Comparación de baremos. El aumento de los puntajes directos a través del tiempo. *Evaluar*, 2(2), 39-51.

Rossi-Casé, L., Neer, R. y Lopetegui, S. (2002). Test de Matrices Progresivas de Raven: Construcción de Baremos y Constatación del "Efecto Flynn". *Orientación y Sociedad*, 3, 181-187.

Rossi-Casé, L., Neer, R. y Lopetegui, S. (2011). Baremo ciudad de La Plata (2000). Buenos Aires, Argentina. Escala General (MPG). En J. C. Raven (Ed.), *Test de Matrices Progresivas. Carpeta de Evaluación, Escala General* (pp. 19-23). Buenos Aires: Paidós.

Rossi-Casé, L., Neer, R., Lopetegui, S., Doná, S. M., Biganzoli, B. y Garzaniti, R. (2014). Matrices Progresivas de Raven: efecto Flynn y actualización de baremos. *Revista de Psicología*, 23(2), 3-13. <http://dx.doi.org/10.5354/0719-0581.2014.36144>

Rossi-Casé, L., Neer, R., Lopetegui, S., Doná, S. M., Biganzoli, B. y Garzaniti, R. (en prensa). Test de Raven: actualización de baremos en adolescentes argentinos y análisis del Efecto Flynn. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*. ISSN: 1135-3848

Rossi-Casé, L., Doná S.M., Garzaniti, B. (en prensa). Test de Raven: Estudio sobre a confiabilidad del uso de la Escala Paralela en sujetos de la ciudad de La Plata, Argentina. La Plata: 5to Congreso Internacional de Investigación de la Facultad Psicología de Universidad Nacional de La Plata.

Rusell, E. W. (2007). Commentary: The Flynn effect revisited. *Applied Neuropsychology*, 14(4), 262-266. Recuperado de <http://dx.doi.org/10.1080/09084280701719211>

Sundet, J. M., Barlaug, D. G. y Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362. <http://dx.doi.org/10.1016/j.intell.2004.06.004>

Teasdale, T. W. y Owen, D. R. (2007). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, 36, 121-126. Recuperado de <http://dx.doi.org/10.1016/j.intell.2007.01.007>

TABLA 1

Composición de la muestra de tipificación del Test de Raven, año 2015, La Plata (Argentina)

Edad	Sexo		Total de casos
	Mujer	Hombre	
13 años	73	135	208
14 años	68	85	153
Totales del grupo de 13-14 años	141	220	361
15 años	82	99	181
16 años	83	72	155
Totales del grupo de 15-16 años	165	171	336
17 años	90	102	192
18 años	73	105	178
Totales del grupo de 17-18 años	163	207	370
Total de casos por sexo	469	598	1067

Nota. Fuente: Muestra, Elaboración propia basada en la administración del Test de Raven-Escala General, en 2012-2015, a estudiantes de ambos sexos, en La Plata, Argentina ($N=1067$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As.

TABLA 2

Comparación porcentual por rango de edades de la población y la muestra de tipificación del Test de Raven, La Plata (Argentina)

Grupo de edades	Censo poblacional 2010	Muestra 2015
13-14 años	32.44	33.83
15-16 años	32.75	31.49
17-18 años	34.81	34.68
Porcentaje total	100.00	100.00

Nota. Fuente: Población, Censo Nacional de Población 2010. Ministerio del Interior de la República Argentina. Muestra, Elaboración propia basada en la administración del Test de Raven-Escala General, en 2012-2015, a estudiantes de ambos sexos, en La Plata, Argentina ($N=1067$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As.

TABLA 3

Test de Raven. Puntuaciones promedio, desviación estándar y rango según edades, año 2015, La Plata (Argentina)

Grupo de Edades	Promedio	Desviación estándar	Rango de puntajes
13-14 años	44.01	7.60	12 a 60
15-16 años	48.39	6.75	10 a 60
17-18 años	49.25	6.33	12 a 60

Nota: $N = 1067$. Puntuación máxima: 60 puntos.

TABLA 4

Comparación de la puntuaciones directas correspondientes al Percentil 50, años 1964, 2000 y 2015, por rango de edades.

Año	Percentil 50		
	13-14 años	15-16 años	17-18 años
1964	29	41	40
2000	48	49	50
2015	45	50	50

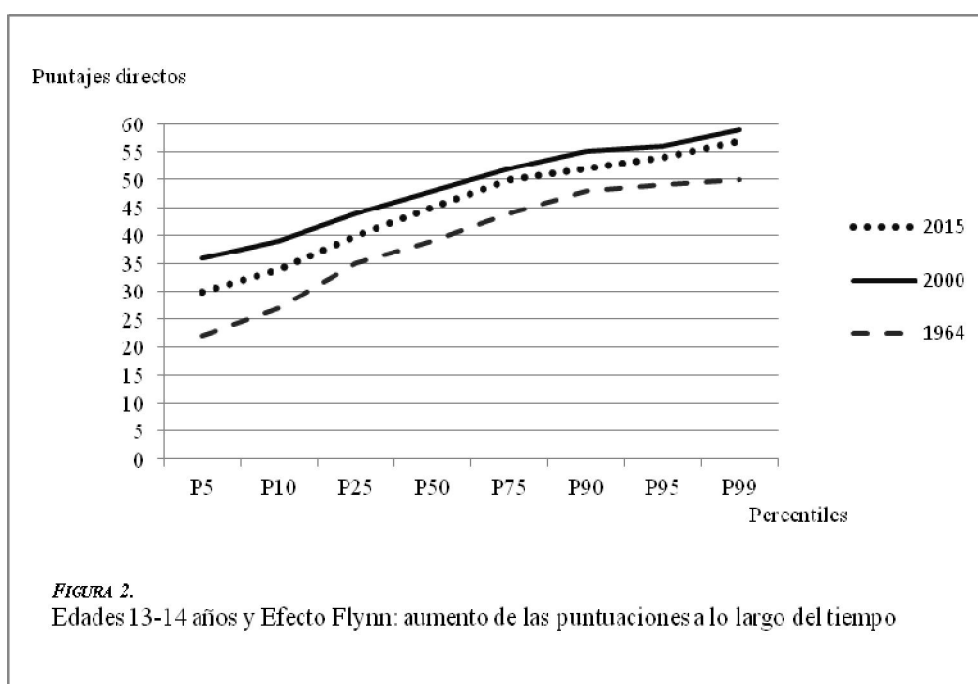
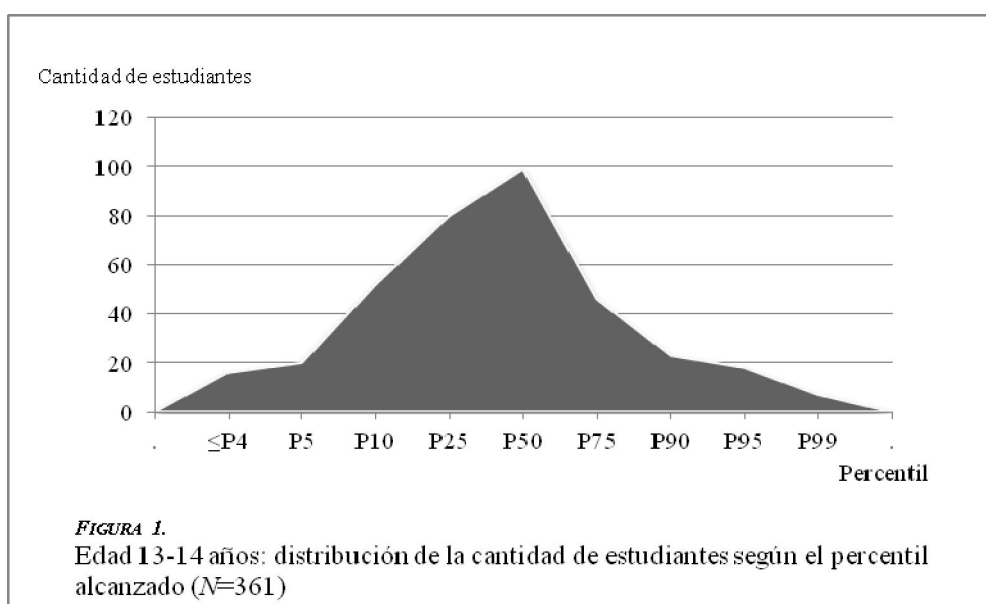
Nota: $N = 1067$. Puntuación máxima: 60 puntos.

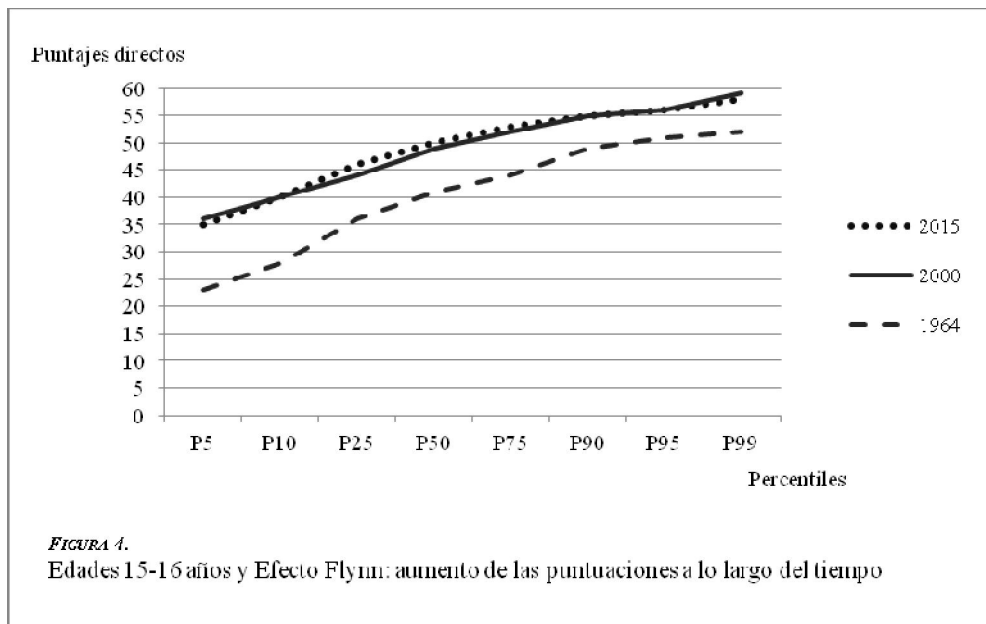
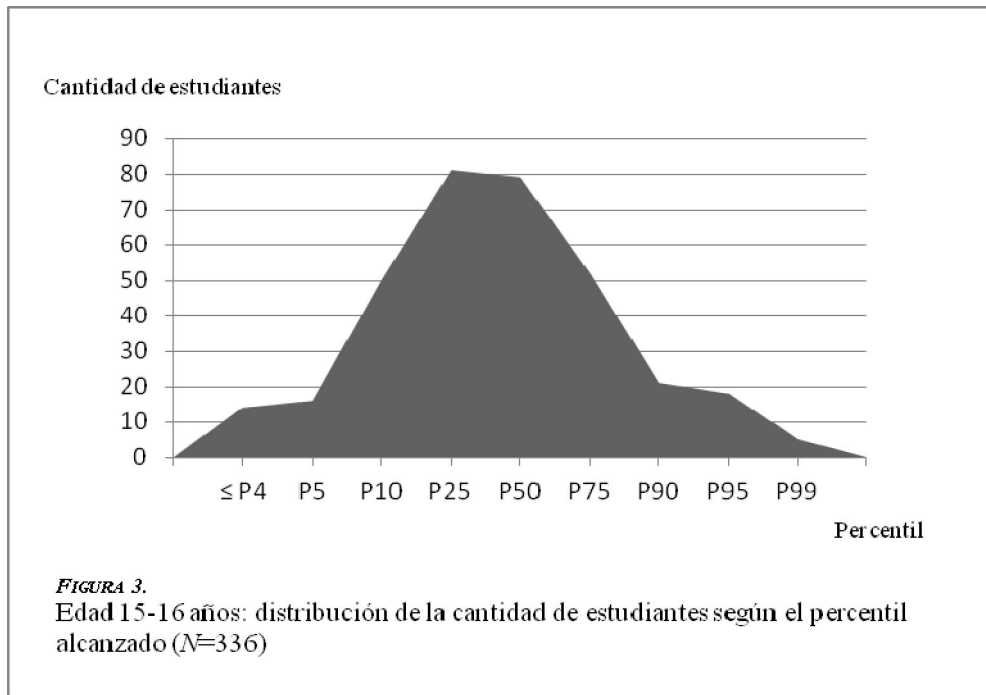
TABLA 5

Promedio de las puntuaciones directas en cada serie del Test de Raven, año 2015, La Plata (Argentina)

Grupo de Edades	Serie A	Serie B	Serie C	Serie D	Serie E
13-14 años	11,30	10,39	8,57	9,02	4,71
15-16 años	11,57	11,01	9,63	9,88	6,31
17-18 años	11,65	10,93	9,76	10,22	6,66

Nota: ($N=1067$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As. Puntuación máxima por Serie: 12 puntos





Cantidad de estudiantes

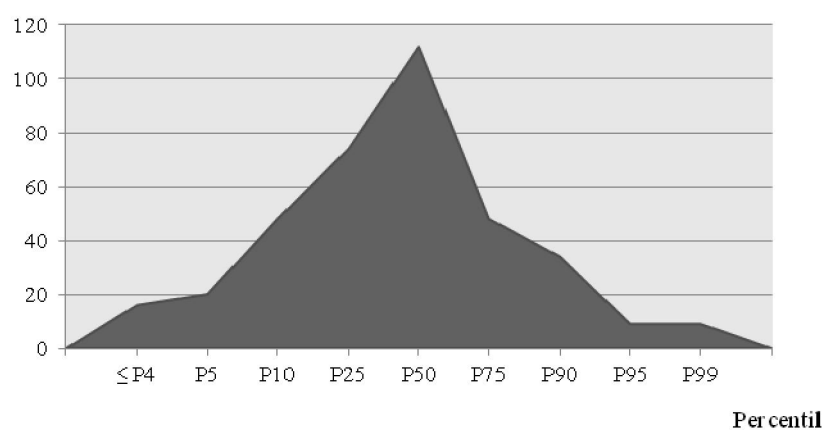


FIGURA 5.

Edad 17-18 años: distribución de la cantidad de estudiantes según el percentil alcanzado ($N=370$)

Puntajes directas

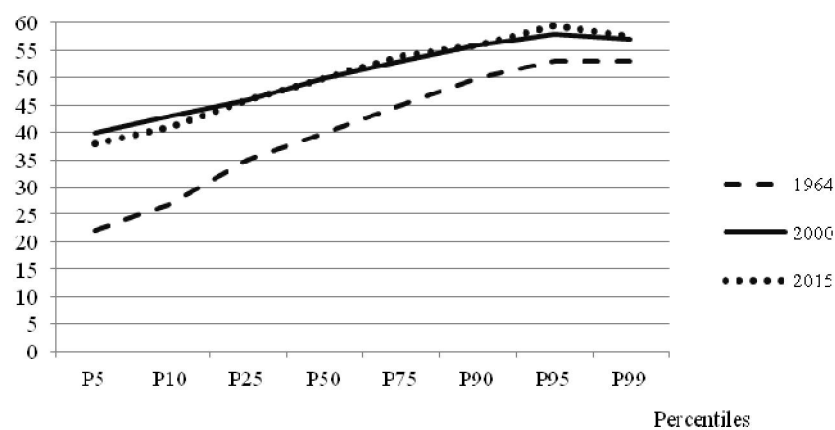


FIGURA 6.

Edades 17-18 años y Efecto Flynn: aumento de las puntuaciones a lo largo del tiempo